

## **IN THE CLAIMS:**

1-19. (Cancelled)

20. (Previously presented) A method for workload balancing in an asynchronous messaging system comprising:

- obtaining an average depth of a message queue;

- controlling a number of server instances of a server for retrieving messages from the message queue based on the average queue depth and one or more predetermined thresholds;

- determining if the average queue depth exceeds a first predetermined threshold of the one or more predetermined thresholds;

- responsive to determining that the average queue depth exceeds the first predetermined threshold, initiating the start of a server instance for retrieving messages from the message queue by placing a trigger message on an initialisation queue, wherein the trigger message indicates that the server instance is to be started; and

- responsive to determining that the first predetermined threshold has been exceeded, resetting the average queue depth to less than the first predetermined threshold immediately upon the start of the server instance.

21-25. (Cancelled)

26. (Previously presented) The method of claim 20, wherein the controlling step comprises:

- terminating the server instance when the average queue depth falls below a second predetermined threshold of the one or more predetermined thresholds.

27. (Previously presented) The method of claim 26 comprising:

- responsive to determining that the average queue depth is below the second predetermined threshold, resetting the average queue depth to greater than the second predetermined threshold immediately upon the termination of the server instance.

28. (Previously presented) The method of claim 26, wherein the step of terminating a server instance comprises at least one of:

- (i) indicating to the server instance that there are no more messages to process on the message queue;
- (ii) indicating to the server instance that a queue manager, controlling the message queue, is shutting down;
- (iii) indicating to the server instance that operator intervention is requesting that the server instance shut down; or
- (iv) requesting that the server instance shut down.

29. (Cancelled)

30. (Previously presented) The method of claim 20, further comprising:  
setting a maximum number of server instances that can be active at any one time.

31. (Previously presented) The method of claim 20, further comprising:  
setting a minimum number of server instances that should be active at any one time.

32. (Previously presented) The method of claim 20, wherein the step of obtaining the average queue depth comprises:  
calculating the average queue depth of the message queue.

33. (Previously presented) The method of claim 32 wherein the step of calculating comprises:  
calculating a time weighted mean average queue depth of the message queue.

34. (Previously presented) The method of claim 32, wherein the step of calculating comprises:  
calculating an exponentially smoothed average queue depth of the message queue.

35-47. (Cancelled)